

**Table S1** General features of *Escherichia coli* genomes sequenced in this work and of *E. coli* reference genomes (as of June 14<sup>th</sup>, 2011).

<i>E. coli</i> strain	Pathotype	Serotype	Genome			Sequencing	
			Size (Mb)	CDS	Accession	Status	Facility
GOS1	EAHEC	O104:H4	5,31	5,217	<sup>a</sup>	171 <sup>b</sup>	Göttingen Genomics Laboratory, Germany
GOS2	EAHEC	O104:H4	5,31	5,224	<sup>a</sup>	204 <sup>b</sup>	Göttingen Genomics Laboratory, Germany
TY-2482	EAHEC	O104:H4	5,29	5,139	<sup>c</sup>	513 <sup>b</sup>	Beijing Genomics Institute, China
LB226692	EAHEC	O104:H4	5,45	5,641	AFOB01000000	364 <sup>b</sup>	Life Technologies, Germany, University of Münster
H112180280	EAHEC	O104:H4	5,31	5,078 <sup>f</sup>	<sup>d</sup>	58 <sup>b</sup> (13 <sup>e</sup> )	Health Protection Agency (HPA), Cambridge, UK
55989 <sup>g</sup>	EAEC	no data	5,15	4,969	CU928145	closed	Genoscope, France
042 <sup>g</sup>	EAEC	O44:H18	5,24	4,810	FN554766	closed	Wellcome Trust Sanger Institute, UK
O157_Sakai <sup>g</sup>	EHEC	O157:H7	5,49	5,363	BA000007	closed	Osaka University, Japan
O103_12009 <sup>g</sup>	EHEC	O103:H2	5,44	5,264	AP010958	closed	University of Tokyo, Japan
H10407 <sup>g</sup>	ETEC	078:K80:H11	5,15	4,746	FN649414	closed	Wellcome Trust Sanger Institute
MG1655 <sup>g</sup>	laboratory	O16*:K12	4,63	4,294	U00096	closed	University of Wisconsin-Madison, USA

<sup>a</sup> <ftp://134.76.70.117>; UserID: EAHEC\_GOS; Password: EAHEC\_GOS

<sup>b</sup> number of contigs bigger than 500 bp

<sup>c</sup> [ftp://ftp.genomics.org.cn/pub/Ecoli\\_TY-2482/Escherichia\\_coli\\_TY-2482.contig.20110606.fa.gz](ftp://ftp.genomics.org.cn/pub/Ecoli_TY-2482/Escherichia_coli_TY-2482.contig.20110606.fa.gz)

<sup>d</sup> <http://www.hpa-bioinformatics.org.uk/lgp/genomes>

<sup>e</sup> number of scaffolds

<sup>f</sup> based on HPA gene prediction

<sup>g</sup> reference genomes used for comparative BiBag analysis

\*O16 = O antigen not expressed in K-12 due to mutation